Some Human Factors Considerations for Designing Mixed Reality Interfaces

Paul Milgram
Dept. of Mechanical and Industrial Engineering
University of Toronto
Toronto, Ontario M5S 3G8
CANADA

ABSTRACT

Mixed Reality (MR) refers to the general case of combining images along a continuum which ranges from purely real (unmodelled) data, such as raw video images, to completely virtual images, based on modelled environments. Depending on where a particular display mode lies on the reality-virtuality continuum, MR encompasses the case of Augmented Reality (AR), as well as the case of Augmented Virtuality (AV). In designing human-machine interfaces for mixed reality applications, a number of considerations are discussed which may potentially impact the effectiveness of the design. In addition to the real-virtual image content (which is closely related to how much knowledge is available about the images being displayed), these include the (visual) perceptual impact of the display technologies used for combining real and virtual images, which manifest themselves in particular when virtual objects must be aligned with real ones, for applications such as AR mediated teleoperation. Other considerations include where the user's particular viewpoint lies along a continuum ranging from ego- to exo-centricity, as well as control-display congruence issues constrained by the other MR factors.

INTRODUCTION

The objectives of this paper are three-fold: a) to review the concept of Mixed Reality (MR) displays; b) to outline a number of human factors considerations which arise as a result of working with MR displays; and c) to propose a taxonomic framework which can be used for distinguishing among a variety of MR display applications.

1. DEFINITION OF MIXED REALITY

The term "Mixed Reality" (MR) has become widely used within the past decade (Milgram & Kishino, 1994; Ohta & Tamura, 1999; Simon & Decollogne, 2006), following its introduction as a method of distinguishing among a variety of display techniques which had previously generally been referred to broadly as "virtual reality" or "virtual environments" or "augmented reality" (e.g. Barfield & Furness, 1995; Azuma, 1997), with apparently little thought given explicitly to the "virtual" and "real" aspect of the related images. With reference to the term "augmented reality", another objective was to extend its definition beyond a relatively limited set of displays.

Milgram, P. (2006) Some Human Factors Considerations for Designing Mixed Reality Interfaces. In *Virtual Media for Military Applications* (pp. KN1-1 – KN1-14). Meeting Proceedings RTO-MP-HFM-136, Keynote 1. Neuilly-sur-Seine, France: RTO. Available from: http://www.rto.nato.int/abstracts.asp.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **01 JUN 2006**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**

4. TITLE AND SUBTITLE
Some Human Factors Considerations for Designing Mixed Reality Interfaces

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Dept. of Mechanical and Industrial Engineering University of Toronto Toronto, Ontario M5S 3G8 CANADA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM002024., The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**UU**	**14**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

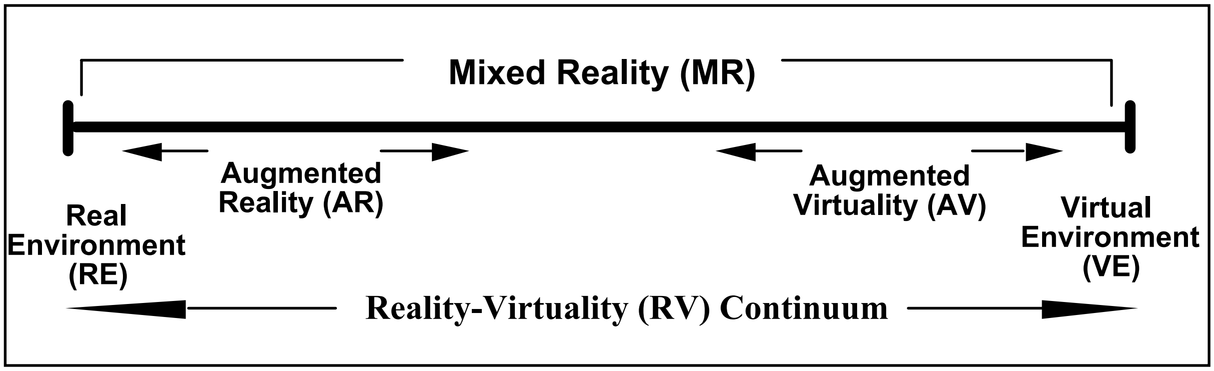

Figure 1: Definition of Mixed Reality (MR) (Milgram & Colquhoun, 1999).

The basic premise underlying MR is that, rather than regarding virtual environments (VEs) and real environments (REs) simply as mutually exclusive opposites, it is helpful to view VEs and REs as opposing poles of a continuous spectrum of possible combinations of real and virtual image content – that is, a *Reality-Virtuality (RV) Continuum*, as illustrated in Figure 1. Using this as our basis, it then becomes rather straightforward to define the term *"Augmented Reality" (AR)*, in a very generic way: an AR image is any representation which involves the augmentation, or enhancement, of a RE based image with some kind of virtual (computer generated) image content. As depicted by the two-sided arrow in Figure 1, AR lies logically at the left side of the RV continuum, with its left border purposely not reaching completely to the real end of the continuum, and its right border intentionally indicating some vaguely defined region in the centre of the continuum. For the sake of equilibrium, it therefore becomes desirable to designate an analogous method of enhancing or extending images arising from purely virtual environments with real image data as constituting *"Augmented Virtuality" (AV)*, depicted analogously in Figure 1.

One direct consequence of this system of definitions is that it significantly widens the number of display methods which can be classified as either AR or AV. For example, one relatively recent definition of AR (presented within the context of *Virtual Environments Standards and Terminology* in the *Handbook of Virtual Environments*) specifies "the use of transparent glasses on which a computer displays data so the viewer can view the data superimposed on real-world scenes" (Blade & Padgett, 2002). Whereas the extent of that definition encompasses what is arguably the most common example of AR, it clearly excludes a significant number of displays which involve superimposing computer generated data onto real-world scenes where the viewer is *not* wearing transparent glasses (for example, "see through video") or where the viewer is simply looking at a monitor. At the opposite end of the MR continuum, our definition of AV allows us to include a wide variety of otherwise strictly VE displays within which some kind of video windows or photographs are included, or where texture mapping onto 3D modelled surfaces is employed (Milgram & Colquhoun, 1999).

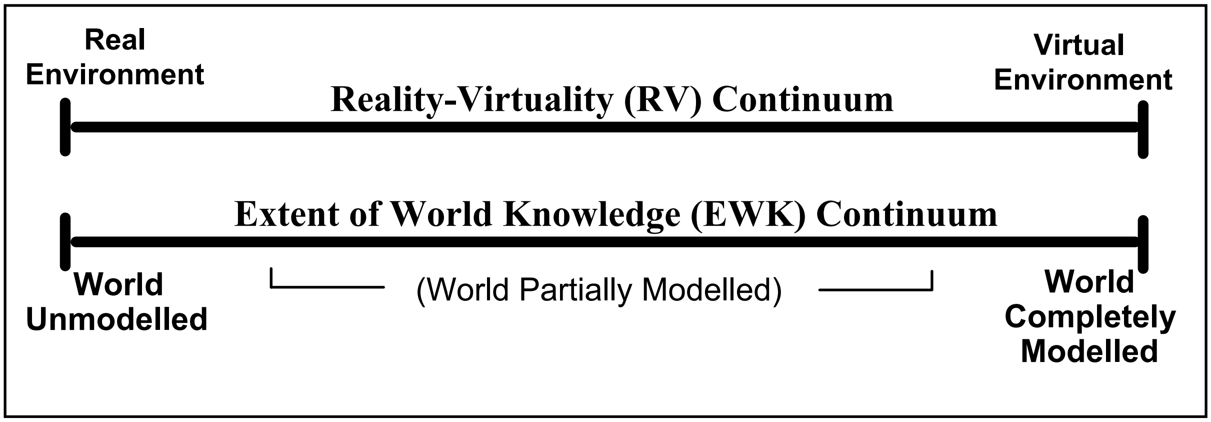

Figure 2: Proposed parallel relationship between *Reality-Virtuality* continuum and *Extent of World Knowledge* continuum. (Milgram & Colquhoun, 1999).

One obviously key element of this classification framework is the definition of what is "real" and what is "virtual". The latter is relatively straightforward; quoting from Blade and Padgett (2002), *virtual* means simply "simulated or artificial", and a *virtual environment* comprises "3D data sets describing an environment based on real-world or abstract objects and data." It is interesting, however, that that same glossary contains no definition of "real" or "real environment". As illustrated in Figure 2, our approach to this challenge is to present a parallel continuum, which we call the *Extent of World Knowledge(EWK) Continuum*, where the knowledge in this case resides collectively within the sensing and computational devices driving the display system in question (Milgram & Colquhoun, 1999). The meaning of the EWK continuum is straightforward; at one end, as in a simple video image, the "system" knows nothing about the data being presented, other than at a pixel level. In other words, even though every pixel (or voxel) in an image has an intensity and colour value, simple sensing of an image imparts no meaning with regards to which object any particular pixel belongs, nor where that object is located within the scene being sensed, nor how that object relates to other objects. In other words, straightforward capturing of a *real environment* image corresponds to an *unmodelled world.* At the opposite end of the spectrum, in contrast, it is easy to recognise that the only way to present the image of a completely virtual environment is if that world is completely modelled. The conclusion is that, even though the two continua are different, they are clearly parallel. Extending this thought further, it therefore becomes useful simply to exploit this parallelism not as a "definition" of real and virtual environments, but as a means of illustrating their differences. Images arising from virtual environments must necessarily be based on completely modelled data, and completely modelled data therefore form a representation of a virtual world (even if that virtual world is a model of a real one). On the other hand, if we know nothing about the environment which we wish to display (an unmodelled world), our sole recourse is to make use of sensed data (i.e. pixel level data) about that world, if such data exist (since otherwise they would be virtual). The logical extension of this line of reasoning, therefore, is that Mixed Reality encompasses images comprising data which are "partially modelled".

2. SOME PERCEPTUAL ISSUES RELATED TO MIXED REALITY

2.1 Interposition conflicts

One extension to the contention that MR encompasses images which are partially modelled is that possessing knowledge about some objects in an image but not others introduces a separate class of problems … as well as

opportunities. Perhaps the most obvious problem which arises is a result of not knowing the identities, dimensions, or locations of objects within images obtained from real world views[1], due to the fact that, consequent to our operational 'definition', everything is unmodelled in such an image. The consequence is that it is difficult to decide which portions of the image must be occluded, in conformity with proper distribution of depth information within the image, with the result being that some objects that are supposed to be farther away end up occluding other objects that are closer … resulting in an obvious perceptual conflict, whereby it can be very difficult to observe an image and know where everything is located. This is exacerbated by the fact that occlusion is arguably the most powerful perceptual cue acting within human depth perception (May & Badcock, 2002; Wickens & Hollands, 2000).

In addition to the problem of not knowing which portions of an image to occlude – that is, by eliminating portions of a graphic object or simulating hidden surfaces in a video image – it is important to realise that often the extent of this issue is a function of the particular display technology being used. To some extent, it is possible to generalise this statement as follows:

- For video based real images, the superimposed virtual graphics almost always occlude the video portions of the image.
- For optically combined images (using, for example, semi-silvered mirrors on a helmet mounted or head–up display), the virtual image portions do not usually completely occlude the real portions, and the real portions do not completely occlude the graphic content; that is, some degree of transparency is usually present.
- For large screen immersive displays, for which the observer's body or tools or furniture (i.e. real image data) interact with either computer generated (virtual) or non-virtual images presented on the display surface, the real image (e.g. the user's hand) always occludes the displayed data.

These occlusion conflict problems are especially well known for the case of video based augmented reality, where the real world comprises video pixels and the virtual image content is computer generated. An important class of applications using such displays includes those which involve the need to align virtual and real objects, for example for making 3D measurements of distances and dimensions within 3D video images (Drascic & Milgram, 1991; Kim et al, 2000), for surgical planning (Dey et al, 2002; Kheddar et al, 2002), as well as those which involve the use of virtual robot images as a vehicle for interactive programming of real robot operations (Kheddar et al, 2002; Milgram et al, 1997).

2.2 Other perceptual issues related to MR technology

In addition to problems caused by the important class of interposition related issues, a number of other issues have been identified, all arising from the fact that the technologies used to sense and present real and virtual images in mixed reality are different. These include the following (Drascic & Milgram, 1996):

- *Luminance limitations and mismatches.* In MR applications involving direct view, the display hardware used can easily result in images that are less bright than direct viewed objects. One consequence of this is that, because brighter objects appear closer, any object that does not result from direct viewing may appear farther away than intended.
- *Contrast mismatches.* Because the contrast ratio of HMDs, monitors and projection systems is typically less than for direct viewing, imaged objects may, once again, appear farther away than appropriate.

[1] Strictly speaking, this statement is not entirely true for the case of 3D mapping images, where we clearly do have information about relative locations. What we do not have, however, is information about to which object each sensed datum belongs … unless some object recognition and/or segmentation algorithms have been applied … in which case the image is no longer completely unmodelled.

- *Resolution and image clarity mismatches.* Directly viewed objects necessarily have more resolution than objects projected using a helmet mounted display (HMD), monitor, or projector, and each of those electronic display systems has its own, usually different, resolution. This can result in different image clarity/fuzziness, mismatches in ocular accommodation, and thus potentially different perceptions of object location (Utsumi et al, 1994).

2.3 Stereoscopic Display Related Technical Issues

Even though the definition of mixed reality given above imposes no requirement that displays be stereoscopic (May & Badcock, 2002; Howard, 2003), it has become quite common for depth information to be added by providing binocular disparity for both real and virtual images (e.g. Dey et al, 2002).[2] Whereas computing appropriate stereo parameters for computer generated virtual images is relatively straightforward, this is not the case for stereoscopic video (SV), which presents such challenges as ensuring proper camera alignment, as well as a suitable field of view and optical magnification, without neglecting potential vertical camera misalignments, optical distortions, video chip misalignments, etc. (Diner & Fender, 1993).

Such issues are further compounded when one attempts to design a stereoscopic mixed reality display system which (by definition) includes both real and virtual image data. Some of the technical issues which one might be expected to encounter include the following (Drascic & Milgram, 1996):

- *Calibration mismatches.* In order for graphic and video images to be scaled properly, the calibration parameters which determine the visual angle, perspective, and binocular parallax of each image (relative to the viewer) must be accurately specified. As mentioned above, achieving this for SV can be non-trivial. The consequence of any calibration mismatch is that graphic and real objects will not match each other when they are supposed to.
- *Dynamic registration mismatches.* Even if one is able to obtain acceptable alignment between real and virtual images, one essentially always faces the ubiquitous challenge of tracking the viewer's head position and orientation, whereby a lag of only tens of milliseconds can cause perceptible mismatches.
- *Interpupillary distance (IPD) mismatches.* In principle, in order for a MR image to appear 'proper', the display should be calibrated not only to the location of the individual viewer's eyes, but also to the IPD, which must be determined ahead of time. Research has shown that even small errors in IPD can lead to large errors in perception of (relative) location of objects in a display.
- *Limited depth resolution.* Because stereo displays rely on presenting disparate left and right eye images, it stands to reason that the depth resolutions obtainable by using different display hardware technologies, each with its own display resolution, will be different.

2.4 Stereoscopic Display Related Perceptual Issues

Whereas the topics outlined above relate primarily to *technological* limitations, there exists an interesting class of issues which are a consequence of some of the perceptual anomalies often encountered when trying to align real and virtual stereoscopic images (Howard, 2003). One of these is illustrated in Figure 3, where two separate problems are illustrated (Drascic & Milgram, 1996). The figure depicts a virtual stereo-graphic (SG) object, the filled square, which is designed to appear floating in space in front of the screen. In order to accomplish this, disparate left- and right-eye images are rendered on the screen surface. When properly designed, the viewer's eyes will converge at the point shown. The first anomaly stems from the fact that, in order to perceive the object described above, the viewer's eyes must converge at one distance, V_{SG}, but must accommodate at another, f_{SG}. Such a conflict causes the viewer to reconcile the two pieces of distance

[2] One obvious exception to this is the case of optical see-through displays, involving natural stereoscopic real world vision, but to which virtual images are added only monoscopically.

information and may result in the conclusion that the object is located at point p, the unfilled square, somewhat closer to the screen than intended by the software designer.

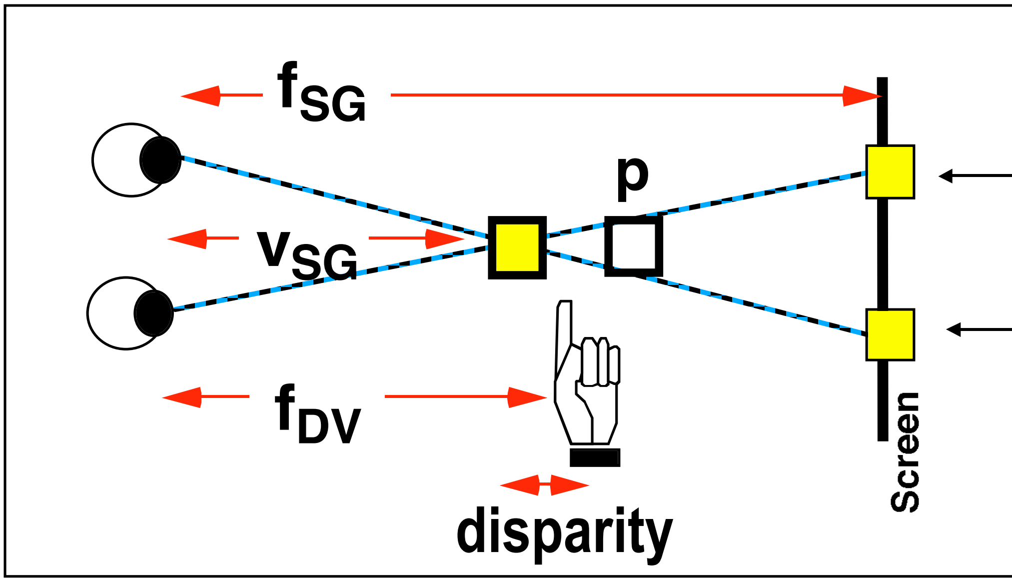

Figure 3: Illustration of accommodation / vergence / interposition interactions.
SG: Stereo-graphics. DV: Direct viewing. f: focus distance. v: vergence distance.

The second anomaly depicted in the figure can result when a real, directly viewed (DV) object, such as the viewer's hand, is placed in proximity to the virtual SG image. If the objective is "to reach out and touch" the virtual object, then clearly both the hand and the object should be at the same location in space. However, in contrast to the accommodation-vergence mismatch encountered with the virtual image, such is not the case for the DV hand, whose focus distance, f_{DV}, and vergence distance, V_{DV} (not shown), are identical. What we have, in other words, is what might be regarded as a 'mismatch of mismatches', which can further affect one's perception of the SG object's location.

Thirdly, although not shown explicitly in the image, we can imagine what might happen as the viewer moves her hand towards the object. If the hand comes between the object and the viewer, then it will occlude the object, causing it to disappear … which is what is supposed to happen. Conversely, if the viewer places her hand between the object and the screen, the hand will still occlude the object, causing it to disappear and concurrently causing an irreconcilable conflict between the combined accommodation and stereoscopic vergence cues and the strong occlusion cue. Such complex interactions can make intended grabbing tasks quite difficult to carry out with such display systems, unless placement accuracy constraints are substantially reduced.

A related conflict occurs whenever one wishes to superimpose a stereoscopic graphic (SG) image on top of, or behind, a stereoscopic video (SV) image of an unmodelled surface (e.g. Dey et al, 2002; Kim et al, 2000). Whenever the SG object is in front of the real surface, everything should appear fine, as shown in Figure 3. However, because graphic images generally occlude real data in such cases, whenever the SG object is *behind* the real SV (occluded) surface, we once again encounter a mismatch … this time between the binocular

disparity cue, which tells the viewer that the object is behind the real surface, and the occlusion cue, which tells the viewer that, it is impossible to continue perceiving an object that is located behind a (non-transparent) surface. In such cases, one of two possible outcomes can occur: the viewer's brain tells her either that the surface must be (semi)transparent – in other words, that both occlusion and binocular disparity cues are compatible – or that, since the surface is (known) not (to be) transparent, the binocular disparity cue must be false – in which case stereoscopic fusion of either the real (SV) surface or the virtual (SG) object breaks down. Research has shown that this effect is strongly influenced by, among other things, the surface texture, or amount of fusible detail, on the surface of the real (SV) surface (Hou, 2002; Hou & Milgram, 2003).

3. SOME VIEWPOINT ISSUES RELATED TO MIXED REALITY

Thus far the discussion has considered essentially only static images; in this section we consider a number of user issues which arise when one wishes to navigate through a mixed reality (MR) world. In general the factors to be discussed here may affect one's ability to perceive spatial relationships and spatial distances, as well as affect one's ability to locomote and manipulate (Loomis & Knapp, 2003; Darken & Peterson, 2002).

In Figure 4 are shown two factors which influence the display frame of reference for either real or virtual environment viewing. For the former case, the camera icon is meant to depict the viewpoint of a real camera which is transmitting an image to a (remote) viewer, whereas for generating VE images the same icon represents a *virtual camera* viewpoint. For both cases the left figure illustrates the effect on a resultant image of camera *attitude* relative to a scene, while the figure on the right illustrates how viewpoint *location* can affect field of view. The direct relationship of this figure to MR environments derives from the simple fact that, for any completely virtual world (which is completely modelled), it is possible to create any image viewpoint desired, thereby enabling complete flexibility, and compatibility with task demands. Conversely, for a completely real environment, of the kind derived for example from an optical sensor or a video camera, the viewpoint presented to the observer must in general correspond to the sensor viewpoint in the real environment.[3] For the latter case, creating an augmented reality (AR) image should not be an insurmountable problem (assuming adequate calibration and registration to the real world), since virtual images can be presented with any pose and any effective viewpoint. For the former case, on the other hand, creating an augmented virtuality (AV) image by adding real sensor data can easily create problems when the real image viewpoint does not match the desired virtual world viewpoint.

[3] It is of course possible to go beyond a single viewpoint, for cases in which, for example, one possesses a database of recorded viewpoints, plus the ability to carry out real-time interpolation among these images. Additionally one might have a large database (potentially 3D), for example a satellite map, which allows interactive panning and zooming through the unmodelled data.

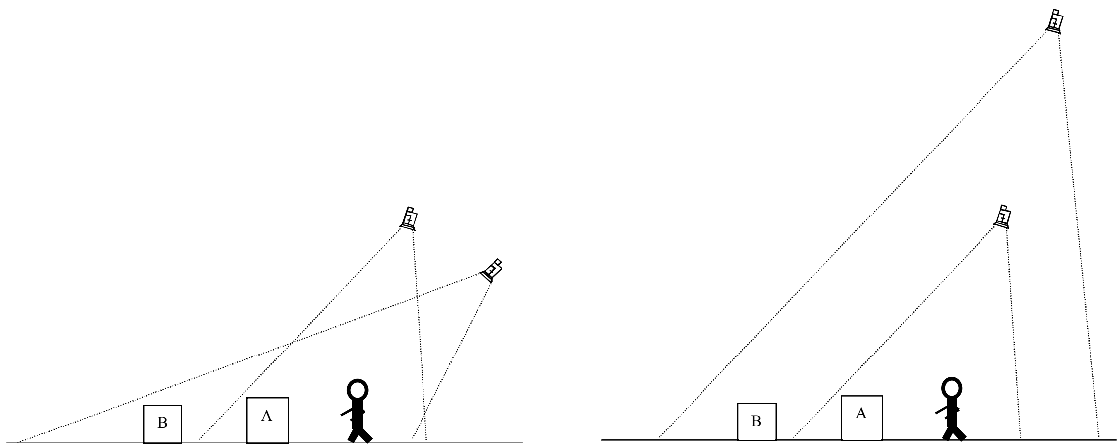

Figure 4: Factors influencing viewpoint in MR (Wang, 2003).

Not depicted explicitly in Figure 4 is how *motion* – either of objects relative to the viewpoint or of the viewpoint relative to the objects, or both – can affect navigation. This particular factor is highly dependent on both location and attitude, in the general sense that, if the (virtual) camera is located relatively far away from the central object of interest – depicted in Figure 4 by an avatar – then the viewing metaphor becomes one of being fixed to the world and watching what transpires as objects / avatars move about within that world – that is, an *exocentric* viewpoint. On the other hand, if the (virtual) camera location is displaced such that it is effectively co-located with the point of view of the central avatar / object of interest, with an attitude corresponding to some nominal viewpoint, then the metaphor becomes one of an *egocentric* viewpoint (Loomis & Knapp, 2003).

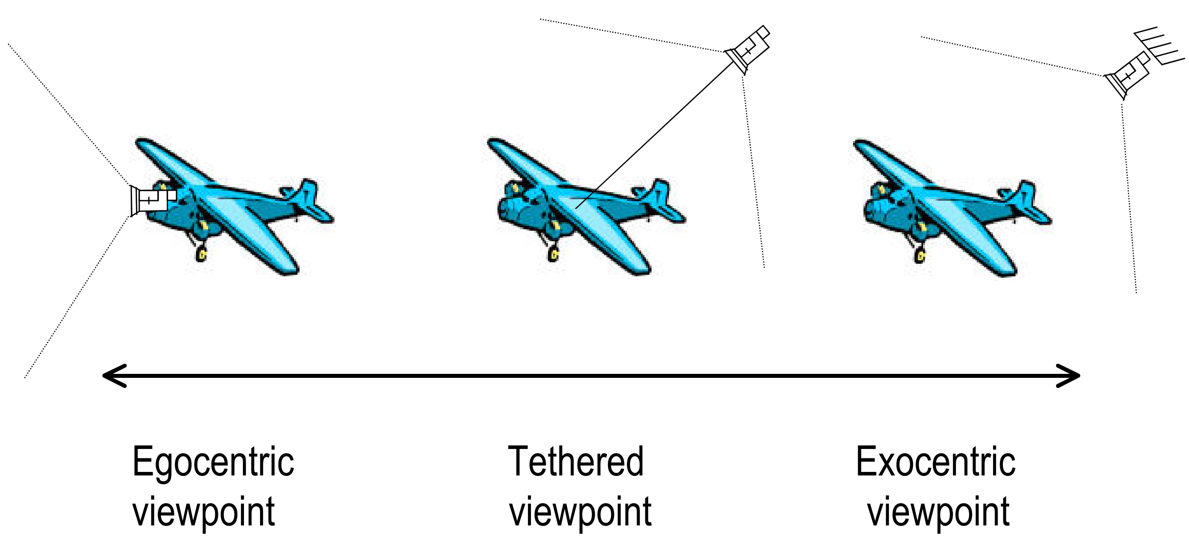

Figure 5: "Centricity continuum," relating egocentric and exocentric viewpoints (Wang, 2003).

This concept of opposing ego- and exocentric viewpoints is illustrated in Figure 5 (Wang, 2003), where the same camera icon is used to represent the effective user viewpoint. What is new in Figure 5, however, is that

we have introduced yet another continuum, this time to depict the idea that there exists a (broad) range of cases between the strictly 'attached' egocentric viewpoint on the left and the strictly 'detached' exocentric viewpoint on the right. In particular, the notion of a *tethered viewpoint* is introduced, depicting the case in which the (virtual) camera viewpoint is towed along behind the central avatar, but at a certain distance (Wickens & Hollands, 2000). One limiting case occurs when the tether length is set to zero, which is then equivalent to the egocentric case shown on the left (assuming that the camera attitude is appropriately aligned). The converse case, for which the tether is made longer, in some ways resembles an approach to exocentric viewing, in the sense that one receives an increasingly larger field of view within which one can observe the motions of one's own avatar.[4]

The significance of the centricity continuum presented in Figure 5 rests on the requirements of the particular task to be carried out with a MR display (Wickens & Hollands, 2000). In general, one can say that egocentric viewing is useful for local guidance and control – which is difficult to acquire with a global "bird's eye" viewpoint. Conversely, exocentric viewpoints are useful for global navigation – which is difficult to acquire with an egocentric "out-the-window" perspective. The theoretical advantage of a tethered viewpoint, in contrast, is that, due to its "intermediate" status, it offers advantages of both ego- and exo-centric viewpoints. Furthermore, if one goes beyond using a tether which is completely rigid, but rather introduces some damping and elasticity – that is, a *dynamic tether* – it has been shown that, for at least some tasks, there exists a set of optimal values of tether rigidity, damping and length, for which both local and global measures of task performance are maximised (Wang, 2003; Wang & Milgram, 2003a, b).

Concluding this discussion of viewpoint centricity, the claim is made that task performance with MR displays will not only depend on the perceptual factors outlined in section 2, which are a function of technological constraints introduced through the combining of real and virtual images, but will also depend on the viewpoints that are made available, within the real-virtual viewpoint constraints mentioned above. We therefore present Figure 6, which depicts a 2D 'design space' within which it is worthwhile to specify where one's MR application lies.

[4] The fact that the camera remains attached to the avatar, rather than fixed to the world, excludes the limiting case of an infinitely long tether corresponding to a purely exocentric viewpoint. In addition, such a limiting tethering case would have to realise a jointly increasing field of view coupled to a continually increasing optical magnification, in order to simulate exocentric viewing at a finite distance.

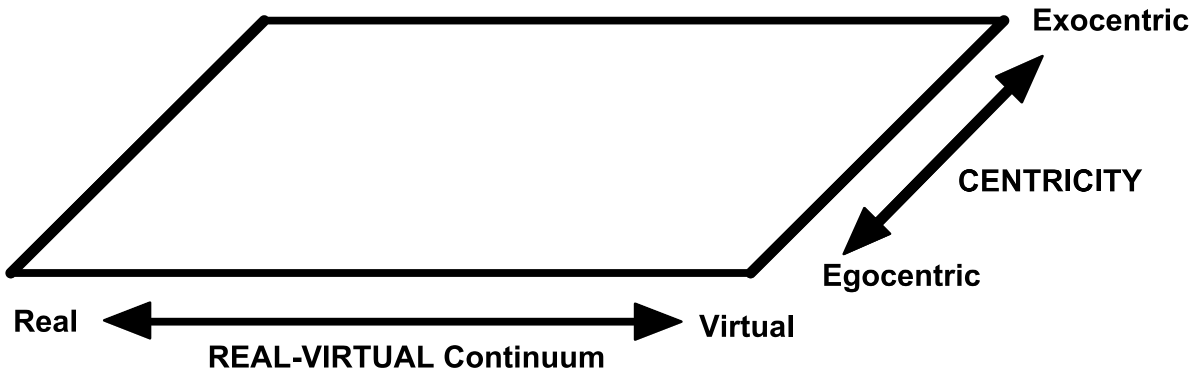

Figure 6: 2D space of MR design tradeoffs: Real-virtual continuum versus Centricity continuum.

4. SOME CONTROL-DISPLAY ISSUES RELATED TO MIXED REALITY

Going beyond merely navigating through mixed reality worlds, we now consider some of the challenges associated with trying to perform manipulations using MR displays. Closely related to the centricity axis in Figure 5 is consideration of the compatibility between available MR display information and the tools provided for effecting operations. The main reason why these factors are related is that there is a corresponding continuum for manipulation, this time relating whether one's actions are referenced to one's own (egocentric) framework – that is, *ego-referenced* control – versus whether one's control inputs must be related to other objects in the scene – that is, *allo-centric* (Klatzky, 1998)[5]. This distinction becomes particularly relevant for operations such as real-time teleoperated control of a manipulator or remote vehicle, for which the particular camera viewpoint will determine whether or not mental rotations, or control transformations, will be necessary (Kheddar et al, 2002). A similar, often more challenging, problem occurs in endoscopic surgery, where the surgeon must manipulate a set of instruments located between herself and the patient, while making use of a real-time camera image whose viewpoint is significantly displaced from the surgeon's natural hand-body space, and which is being controlled by another person (Satava & Jones, 2002). Figure 7 illustrates the control frame of reference issue by the continuum labelled *Control/Display (C/D) Alignment*. The message there is that complete compatibility, or high C/D alignment, analogous to normal motor actions using one's own hands and one's own eyes, exists on the left side; however, as more viewpoint displacements are added, the C/D offset increases, and the congruence between controls and displays becomes greater.

Two other continua are depicted in Figure 7. The first of these, relating *Direct Control* to *Indirect Control*, refers to, on the left, whether the operator is able to use her own hands / limbs, or the equivalent of her own hands / limbs – that is, isomorphism – versus increasing complexity of tool use, towards the right. The remaining continuum refers to the Control Order, with the straightforward message being that higher order control (of tools) is more complex, and thus contributes to decreasing congruence. This leads us to the principal message of Figure 7, that there exists a Control-Display Congruence Continuum, influenced collectively by a number of different factors, including the three shown in the figure, whose effect is to influence the ease and efficiency with which an operator will be able to carry out remote operations.

[5] For the sake of simplicity, it is possible to think of this distinction in terms of *ego-* and *world-referenced* control frameworks.

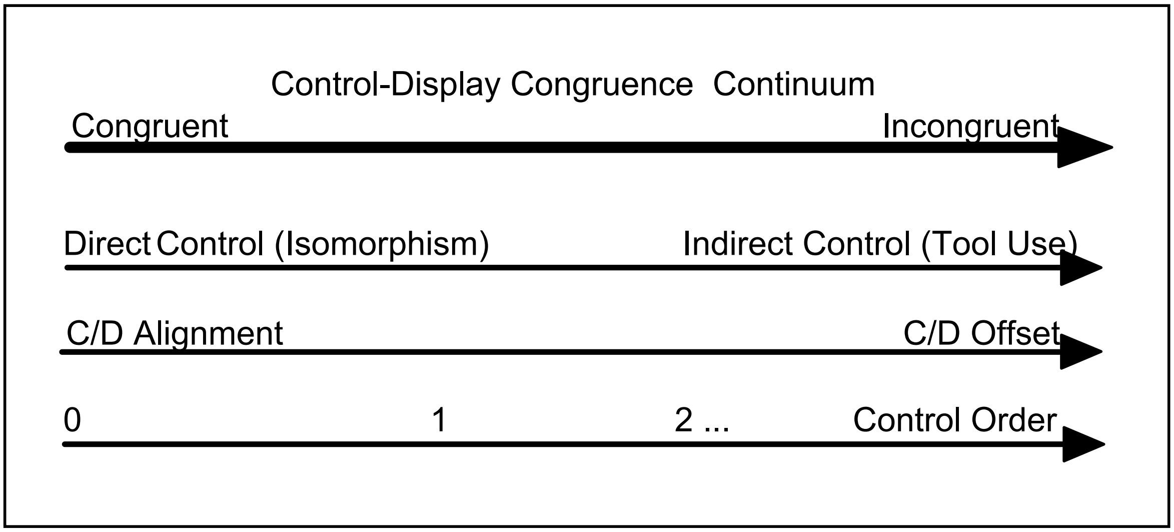

Figure 7: Factors contributing to control-display congruence. (Milgram & Colquhoun, 1999).

The relationship of this continuum to the centricity continuum has already been discussed. With respect to the reality-virtuality continuum, we observe (again) that, in a completely virtual environment, one can ordinarily easily adjust one's viewpoint to bring about reasonably high control-display congruence. In a completely real environment, on the other hand, one does not typically have this ability, resulting in potentially low control-display congruence. Mixed reality environments therefore present the challenge of matching virtual to real, as well as the potential means of overcoming some of the challenges of C/D incongruence.

5. CONCLUSION: TAXONOMY OF MIXED REALITY APPLICATIONS

To conclude this discussion, the three dimensional framework of mixed reality applications is presented in Figure 8 (Milgram & Colquhoun, 1999), in relation to which the messages deriving from the present discussion can be summarised as follows:

- When designing a MR application, it is important to keep in mind the extent of knowledge which is available about the real and virtual elements of the display, since this will influence factors such as the flexibility of viewpoint manipulation.
- The relationship between real and virtual elements of a MR image can potentially have a significant impact on how aspects such as perception of relative spatial object locations are perceived within the image.
- When designing a MR application it is potentially important to be able to trade off local guidance and control (compatible with egocentric viewing) against global spatial awareness (compatible with exocentric viewing). The possibility may exist to transit between these extremes using techniques such as dynamic viewpoint tethering, thereby maintaining visual momentum across transitions.
- Control-display congruence is (obviously) preferable to C/D incongruence; however, whether this is achievable may be highly dependent on the particular RV mixture, and one's flexibility with respect to the centricity continuum.

- Mixed reality encompasses a wide variety of interactive applications. In one sense it is useful to have a single label to which several applications can be related. More useful, however, is the potential for specifying the *distinctions* among different areas of research and development. It is proposed that situating particular application within the taxonomic framework of Figure 8 can be a useful means for helping practitioners specify the commonalities, and differences, between their various endeavours.

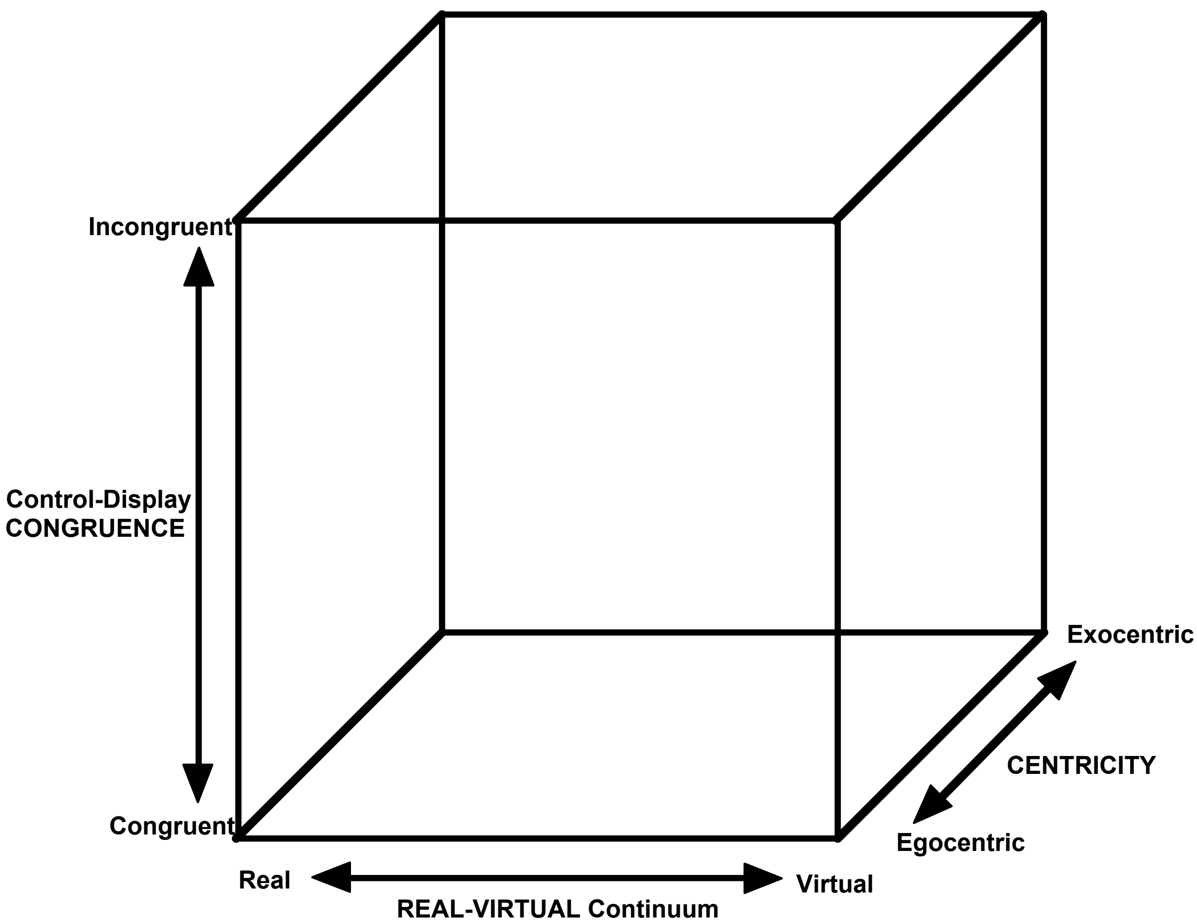

Figure 8: Taxonomy of MR design tradeoffs.

REFERENCES

Azuma RT, "A survey of augmented reality" *Presence*, 6(4), 355-385, 1997.

Barfield W & Furness III TA (ed's). *Virtual Environments and Advanced Interface Design.* Oxford University Press, New York 1995.

Blade RA & Padgett ML, "Virtual environments standards and terminology." Chapter 2 (15-27) in KM Stanney (ed), *Handbook of Virtual Environments: Design, Implementation, and Applications*. Erlbaum Associates, New Jersey 2002.

Darken RP & Peterson B. "Spatial orientation, wayfinding, and representation." Chapter 24 (493-518) in KM Stanney (ed), *Handbook of Virtual Environments: Design, Implementation, and Applications*. Erlbaum Associates, New Jersey 2002.

Dey D, Gobbi DG, Slomka PJ, Surry KJ, & Peters TM, "Automatic fusion of freehand endoscopic brain images to three-dimensional surfaces: Creating stereoscopic panoramas". *IEEE Trans Med Imaging*, 21(1): 23-30, 2002.

Diner DB & Fender DH. *Human Engineering in Stereoscopic Viewing Devices*. Plenum Press, NY & London, 1993.

Drascic D & Milgram P, "Positioning accuracy of a virtual stereoscopic pointer in a real stereoscopic video world", *Proc. SPIE Vol. 1457, Stereoscopic Displays and Applications II*, February 1991.

Drascic D & Milgram P, "Perceptual issues in augmented reality", *Proc. SPIE, Vol. 2653, Stereoscopic Displays and Virtual Reality Systems III*, Jan.-Feb. 1996.

Hou M. *Effects of surface characteristics on alignment of real and graphic objects in stereoscopic augmented reality environments.* PhD dissertation, University of Toronto, 2002.

Hou M & Milgram P. "A sensitivity study of factors influencing real-virtual object alignment performance in stereoscopic augmented reality environments". *Proc. 47st Annual Meeting of Human Factors & Ergonomics Society*. Denver, Colorado, 1630-1634, Oct. 2003.

Howard IP. "Fidelity of disparity-based stereopsis" Chapter 11 (219-245) in Hettinger LJ & Hass MW (eds) *Virtual and Adaptive Environments: Applications, Implications, and Human Performance Issues.* Erlbaum, New Jersey 2003.

Kheddar A, Chellali R & Coiffet P, "Virtual environment-assisted teleoperation." Chapter 48 (959-997) in KM Stanney (ed), *Handbook of Virtual Environments: Design, Implementation, and Applications*. Erlbaum Associates, New Jersey 2002.

Kim M, Drake JM & Milgram P, "Virtual tape measure for the operating microscope: System specifications and performance evaluation", *Computer Aided Surgery,* 5:148-155, 2000.

Klatzky RL. "Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections" pp 1-17 in C Freksa, C Habel & KF Wender (eds), *Spatial Cognition –An Interdisciplinary Approach to Representation and Processing of Spatial Knowledge*. Springer-Verlag Berlin 1998.

Loomis JM & Knapp JM. "Visual perception of egocentric distance in real and virtual environments." Chapter 2 (21-46) in Hettinger LJ & Hass MW (eds) *Virtual and Adaptive Environments: Applications, Implications, and Human Performance Issues.* Erlbaum, New Jersey 2003.

May JG & Badcock DR, "Vision and virtual environments." Chapter 3 (29-63) in KM Stanney (ed), *Handbook of Virtual Environments: Design, Implementation, and Applications*. Erlbaum Associates, New Jersey 2002.

Milgram P & Kishino F, "A taxonomy of Mixed Reality visual displays*", IEICE Trans. on Information and Systems* Vol. E77-D, No. 12, pp. 1321-1329, Dec. 1994.

Milgram P & Colquhoun H, "A Taxonomy of Real and Virtual World Display Integration." Chapter 1 (1-16) in Y Ohta and H Tamura (Ed's), *Mixed Reality – Merging Real and Virtual Worlds.* Ohmsha (Tokyo) & Springer Verlag (Berlin), 1999.

Milgram P, Yin S & Grodski JJ, "An augmented reality based teleoperation interface for unstructured environments." *Proc Am. Nuclear Sociey 7th Topical Mtg on Robotics & Remote Systems,* 966-973, 1997.

Ohta Y & Tamura H (Ed's), *Mixed Reality – Merging Real and Virtual Worlds.* Ohmsha (Tokyo) & Springer Verlag (Berlin), 1999.

Satava RM & Jones SB. "Medical applications of virtual environments." Chapter 47 (937-957) in KM Stanney (ed), *Handbook of Virtual Environments: Design, Implementation, and Applications*. Erlbaum Associates, New Jersey 2002.

Simon G & Decollogne J, *Intégrer Images Réelles et Images 3D: Post-production et réalité augmentée.* Dunod, Paris 2006.

Utsumi A, Milgram P, Takemura H & Kishino F. "Investigation of errors in perception of stereoscopically presented virtual object locations in real display space", *38th Annual Meeting of Human Factors and Ergonomics Society*, Boston, October 1994.

Wang W. *Human navigation performance using 6 degree of freedom dynamic viewpoint tethering in virtual environments.* PhD dissertation, University of Toronto, 2003.

Wang W & Milgram P. "Viewpoint optimisation for virtual environment navigation using dynamic tethering - A study of tether rigidity". *Proc. 47st Annual Mtg Human Factors & Ergonomics Society*. Denver, Colorado, 2128-2132, Oct. 2003a.

Wang W & Milgram P. "Effects of viewpoint displacement on navigational performance in virtual environments". *Proc. 47st Annual Mtg Human Factors & Ergonomics Society*. Denver, 139-143, Oct. 2003b.

Wickens CD & Hollands JG, *Engineering Psychology and Human Performance.* (3rd edition). Prentic-Hall, New Jersey, 2000.